

VOLUME 66
NUMBER 2

WHOLE No. 334
1952

Psychological Monographs: General and Applied

Combining the *Applied Psychology Monographs* and the *Archives of Psychology*
with the *Psychological Monographs*

HERBERT S. CONRAD, *Editor*

Reliability and Validity of Judges' Ratings of Adjustment on the Rorschach

By

MARGUERITE Q. GRANT, VIRGINIA IVES,
and JANE H. RANZONI

Institute of Child Welfare, University of California

Accepted for publication March 28, 1951

Price \$1.00

Published by

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

1515 MASSACHUSETTS AVE. N.W., WASHINGTON 5, D.C.

Psychological Monographs:
General and Applied

HERBERT A. SIMON

COPYRIGHT, 1952, BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

Ratings of Adjustment on the
Rorschach

MARGUERITE G. CHANT, VIRGINIA LEE
and JANE H. RANSOME

Department of Psychology, University of California

THE UNIVERSITY OF CALIFORNIA PRESS
400 UNIVERSITY AVENUE, BERKELEY, CALIF. 94720

I. INTRODUCTION¹

THIS research was designed to answer three main questions which grew out of previous Rorschach research in the Guidance Study at the Institute of Child Welfare, University of California. These questions are: (1) Can Rorschach "experts" agree on ratings of adjustment based on the Rorschach? (2) In what way do increasing amounts of Rorschach material affect such judgments? (3) If substantial inter-rater agreement can be demonstrated, what relationship can be shown with outside criteria?

In a previous article from the Institute of Child Welfare (5), the possibilities and limitations inherent in the "sign" approach to the Rorschachs of a normal adolescent population have been explored. The article presented data indicating that the incidence of the so-called "neurotic" signs in this population was so high as to make meaningless the label "neurotic." The possible hypothesis that adolescence is a time of emotional turmoil and that one could, therefore, expect many neurotic signs, was not supported by the data. With this hypothesis one would predict a decrease in the number of signs as the crisis period in ado-

lescence began to merge into adulthood. In our group, no consistent decline was evident over the age range 11 to 18 years. Conversely, the indicators of normality (Rorschach "adjustment" signs) were low; their correlations with interview ratings of adjustment varied nonsignificantly around zero.

The results of that study made it apparent that the "sign" method offered little in a systematic and fruitful approach to the understanding of the Rorschachs of our cross section of adolescents. These findings may not be surprising to many Rorschach workers. One of the major characteristics claimed for the test has been that, for a meaningful understanding of the dynamics of personality organization, the Rorschach must be used as a gestalt. Clinicians agree that the more atomistic the Rorschach analysis, the less its worth as an evaluative instrument. With this in mind, for the present study, three experienced Rorschach workers were asked to make subjective judgments of adjustment on the basis of the Rorschach records.

II. THE SAMPLE AND CRITERION

The subjects were 71 boys and 75 girls who were members of a longitudinal growth study and had been randomly selected 18 years previously from the birth registry in Berkeley, California.² The Rorschach had been administered as part of the testing program of the study, beginning when the children were 11 years of age. As a result, the 18-year records used in this experiment are the fourth Rorschach for most of the subjects. These records were chosen for three

²For a complete description of the sample, see Macfarlane (8). For a discussion of Rorschach procedures in the Guidance Study at the Institute, see McFate and Orr (9).

¹Clerical work and statistical analyses were made available by a grant-in-aid from the U. S. Public Health Service. The study proper was made possible by assistance from the Laura Spelman Fund, the General Education Board, and the Rockefeller Foundation.

Special acknowledgement is due to Jean W. Macfarlane, Director of the Guidance Study, Institute of Child Welfare, for her help and suggestions throughout this study; to Lucile Allen for making special adjustment ratings of the C group; to Helen Clementson for acting as one of the judges; to Edith Katten and Arline Burton for their roles as control raters; and to Frances G. Orr and Margaret Snyder for their work in administering the Rorschachs. The authors also wish to thank Dr. Rheem Jarrett and Charles Hanley for their statistical advice, and Brittomar Handlon for assistance in computation.

reasons. First, they are the most "adult" records and, therefore, more comparable to other "adult" norms in the literature. Second, age 18 has the largest number of cases at any one age in this sample, and, third, these records are longer and richer than earlier tests and thus offer a better basis for rating.

The outside criteria of adjustment consist of ratings of current adjustment made by case workers and based on considerable interview contact with the subject over a period of years. The areas of self-adjustment, social adjustment, and family adjustment were among the variables rated. These ratings and the interviews upon which they were based will be described in detail in the section on validity.

III. PROCEDURE

A. JUDGES

The ratings of the Rorschach protocols were made by three women judges. Since ratings depend upon the training and Rorschach experience of raters and the breadth of the sample upon which the experience is based, a brief statement about each rater follows.

Judge A was trained by Bruno Klopfer in 1939 and has subsequently attended several "workshops." In the course of ten years of administering and teaching the Rorschach, she has seen and evaluated well over 5,000 records, including all ages and both normal and pathological cases.

Judge B was initially trained by Judge A in 1943 and has subsequently had several "workshops" with Bruno Klopfer. She has worked with other psychologists, well trained in the Rorschach, both in the clinic and in academic settings, and has spent two years teaching and supervising Rorschach work in a clinic.

Judge C attended a Hertz "workshop" in 1946 and has since used the test in a well-established clinic under expert supervision.

For purposes of control, two other raters were used in part of the experiment. The Rorschach experience of these

"nonexpert" raters is described in Section IV, Part 4.

B. THE RORSCHACH RATING SCALE

A four-point scale was used for sorting the Rorschach: (I) very maladjusted, (II) moderately maladjusted, (III) fairly well adjusted, and (IV) very well adjusted. These scale points were not further defined.

C. THE SORTING PLAN

The records were divided into four groups for the purpose of rating: 41 Guidance (G) group boys; 37 Guidance (G) group girls; 30 Control (C) group boys; and 38 Control (C) group girls.³ Three sorts into the four adjustment categories were made on each group of records, with successively increasing amounts of Rorschach information given to the judges on each. This plan was followed in order to determine the contributions of both the formal aspects and the content to the reliability of the ratings and the capacity of the tool to predict clinical judgments.

For Sort 1, the only Rorschach material on each case consisted of the summary page of the standard Klopfer and Davidson scoring form, including the determinant profile and the standard percentages and ratios. In addition, the raters were given the norms of this particular 18-year-old sample, in the form of means, medians, quartiles, and per cent occurrence for each of the four groups.

For Sort 2, the judges were given the previously mentioned summary page plus the tabulation page of the scoring form, which shows the distribution of the

³ The primary difference between the G (Guidance) and C (Control) groups is the intensity with which they have been studied through interviews. The Rorschach administrations were equivalent for the two groups.

scored variables over the ten cards. On Sort 3, the raters were given the entire record, including the full protocol.

The protocols were presented in a random order, with this order changed from sort to sort. Any case-identifying material was deleted, since one of the judges had also acted as an interviewer with some of the children whose Rorschachs were used, but the sex of the subject was known. On each sort, both boys and girls in the *G* or *C* groups were rated before continuing to the next sort, so that one to two days elapsed between the rating and rerating of the same Rorschach records. The judges did not have access to their previous ratings. This fact, plus the size of the groups sorted, probably minimized the memory factor in serial rating of the same records.

The instructions to the judges were (a) to take into consideration both self- and social adjustment in their ratings, (b) to strive for a normal distribution of ratings over the four-point scale, (c) to try to rate adjustment in terms of the norms of *this* group, and (d) to record separately the cues used for individual ratings when this was possible.

After the three sorts were completed on all groups, the judges conferred in an attempt to make explicit the cues they had used in their ratings and discussed specific cases in which disagreement occurred.

IV. RESULTS AND DISCUSSION

A. RATER BEHAVIOR

1. Distributions of ratings over three sorts

All of the judges found it difficult to approximate a normal distribution on the four-point adjustment scale even though they had been asked to attempt it and to rate in terms of this group. The distributions of the cases remained quite constant as more Rorschach information was given to the raters. (See Table 1. The distribution of cases was so similar for all judges that the results of the three judges' ratings are averaged for this table.) On all three sorts for both boys and girls, 61 to 71 per cent of the cases were rated as "maladjusted" (Ratings I and II). With the exception of the first sort of the boys' group, the differences in the per cent of cases rated as "maladjusted" and those rated as "adjusted"

TABLE 1
PERCENTAGE DISTRIBUTION OF RATINGS OVER THREE SORTS

Adjustment Category	Boys (<i>N</i> = 71)*			Adjustment Category	Girls (<i>N</i> = 75)*		
	Sort				Sort		
	I	2	3		I	2	3
I	21%	27%	30%	I	16%	16%	17%
II	40%	44%	41%	II	50%	49%	50%
III	31%	25%	27%	III	26%	30%	26%
IV	8%	4%	2%	IV	8%	5%	7%
	100%	100%	100%		100%	100%	100%

Sort 1—Profile page only

Sort 2—Profile page plus tabulation sheet

Sort 3—Entire record including content

Adj. category I—very maladjusted

Adj. category II—moderately maladjusted

Adj. category III—fairly well adjusted

Adj. category IV—very well adjusted

* Sorts were done in four groups: *G* boys, *N* = 41; *C* boys, *N* = 30; *G* girls, *N* = 37; *C* girls, *N* = 38.

are statistically significant above the 5 per cent level of confidence. Many more cases were rated as "very maladjusted" than were judged "very well adjusted" by all three raters. A larger percentage of boys than of girls were rated "very maladjusted" by all judges. There is some tendency for the percentage of boys placed in adjustment Category I to increase as more Rorschach information is given the judges. A corresponding decrease occurs in the number of cases placed in Category IV. This change does not take place in the girls' groups.

2. Intra-rater agreement

As more Rorschach material was made available over the three sorts, the judges tended to change the adjustment rating of approximately half of the cases by at least one category (see Table 2), in spite of the fact that the percentage of cases in each adjustment group remained quite constant. It might have been expected that the greatest number of changes in rating would occur between Sorts 1 and 3. This was true in the case of the girls' groups, but not consistently in the boys'. Other sex differences are also evident. In the ratings of all judges, the least change for the girls' groups occurs between Sorts 1 and 2, but for the

boys' groups, between 2 and 3. These sex differences seem in line with many other findings in the Guidance Study Rorschach studies. That is, the content and verbalizations (introduced in Sort 3) of the girls' records are much richer than those of the boys', and so might tend to influence to a greater extent the change in the Rorschach adjustment ratings given to the girls.

3. Inter-rater agreement

Table 3 presents an affirmative answer to one of the questions raised in the introduction, namely, whether experts can agree on over-all adjustment ratings made from Rorschach protocols. Although far from unity, the tetrachoric correlations of inter-rater agreement all show an acceptable degree of relationship between the ratings of the three judges.

The sex difference is again apparent. In rating the boys' records, there is an increase in inter-judge agreement from the first to the second sort and a slight decrease from the second to the third sort as more Rorschach information is given. The correlations decrease slightly on the second sort for the girls, then increase on the third.

A comparison of the full and partial

TABLE 2
INTRA-RATER AGREEMENT:
PERCENTAGE OF JUDGES' SELF-AGREEMENT IN RATINGS OVER THREE SORTS

Per Cent of Cases Constant Between:	Boys (N = 71)			Per Cent of Cases Constant Between:	Girls (N = 75)		
	Rater				Rater		
	A	B	C		A	B	C
Sorts 1 and 2	61%	51%	54%	Sorts 1 and 2	73%	61%	60%
Sorts 2 and 3	63%	59%	61%	Sorts 2 and 3	61%	48%	57%
Sorts 1 and 3	56%	54%	54%	Sorts 1 and 3	45%	43%	49%

Sort 1—Profile page only

Sort 2—Profile page plus tabulation sheet

Sort 3—Entire record including content

TABLE 3
CORRELATIONS OF INTER-RATER AGREEMENT ON ADJUSTMENT RATINGS*

Raters	Boys (N = 71)			Raters	Girls (N = 75)		
	Sort				Sort		
	1	2	3		1	2	3
A×B	<u>.61</u>	<u>.85</u>	<u>.68</u>	A×B	<u>.70</u>	<u>.65</u>	<u>.72</u>
A×C	<u>.67</u>	<u>.71</u>	<u>.70</u>	A×C	<u>.45</u>	<u>.42</u>	<u>.56</u>
B×C	<u>.61</u>	<u>.75</u>	<u>.72</u>	B×C	<u>.73</u>	<u>.50</u>	<u>.60</u>

* Significance of the tetrachoric correlations is indicated as follows:

Significantly different from zero at 1 per cent level,
Significantly different from zero at 5 per cent level,

agreement of the judges' ratings is shown in Table 4. (Since all of the distributions were so skewed, these are probably a better measure than the tetrachoric correlations.) Over the three sorts, on 29 per cent of the girls and 34 to 39 per cent of the boys, the three judges agreed completely on the adjustment rating given. When, to the cases of complete agreement, we add those cases in which two judges agreed on a rating category and the third judge disagreed by one

scale point only, 85 to 98 per cent of all cases are included. The other 2 to 15 per cent consist of those cases in which two judges agreed and the third judge disagreed by two scale points or in which the three judges gave three different ratings. In addition to the above-mentioned lower percentage of girls than boys on which the judges completely agreed, Table 4 shows a higher percentage of girls on which the judges completely disagreed.

TABLE 4
INTER-RATER AGREEMENT AND DISAGREEMENT ON THREE SORTS

Extent of Agreement	Boys (N = 71)			Girls (N = 75)		
	Sort			Sort		
	1	2	3	1	2	3
Three judges agree	39%	34%	38%	29%	29%	29%
Two judges agree and third judge disagrees by one scale point	46%	64%	58%	58%	62%	56%
Two judges agree and third judge disagrees by two scale points	8%	1%	0%	4%	4%	3%
Two judges agree and third judge disagrees by three scale points	0%	0%	0%	0%	0%	0%
Three judges disagree	7%	1%	4%	9%	5%	12%
	100%	100%	100%	100%	100%	100%

Sort 1—Profile page only

Sort 2—Profile page plus tabulation sheet

Sort 3—Entire record including content

It was expected that because of similarity of training and experience background, Judges A and B would agree more highly with each other than either would with Judge C. Neither the evidence in Table 3 nor a further breakdown of Table 4 into the agreements and disagreements of individual raters indicates any such tendency. The average per cent agreement of the judges over all sorts is as follows: A with B, 56 per cent; A with C, 51 per cent; B with C, 53 per cent.

4. Sorting by Rorschach "novices"

After discovering the extent to which Rorschach "experts" could agree on adjustment ratings from Rorschach protocols, the authors asked two more individuals to act as judges in rating one of the four groups described in this paper. One of them, Judge D, has had many years of experience interviewing as a social worker. With respect to the Rorschach, she can be considered "naive," since her only experience with the test

was as a subject a year prior to this study. Judge E is a graduate student in clinical psychology, who had, at the time of these ratings, taken a one-semester course in Rorschach administration and scoring from Judge A. Since the formal scoring aspects of the records meant nothing to Judge D and relatively little to Judge E, they were given the entire protocols to rate. Thus, their one sort was comparable to Sort 3 for Judges A, B, and C. Judge D worked with the records of the *C* girls ($N = 38$) and Judge E with the records of the *G* boys ($N = 41$).

Table 5 presents the inter-rater agreement of the Rorschach "novices" with the "experts" and the agreement among the "experts" in the ratings of the groups with which the "novices" worked. Comparing these results, we find that Judge D's correlations with the original judges are only slightly lower than the "experts" agreements among themselves, and that Judge E correlates higher with Judges A, B, and C than they did with each other.

TABLE 5
INTER-RATER AGREEMENT (TETRACHORIC r 's) OF RORSCHACH "NOVICES" WITH "EXPERTS,"
AND OF "EXPERTS" WITH EACH OTHER, ON COMPARABLE SORTS*

Raters	Rater D	Raters	Experts**
	<i>C</i> Girls		<i>C</i> Girls
D×A	.55	A×B	.63
D×B	.33	A×C	.55
D×C	.49	B×C	.57
Raters	Rater E	Raters	Experts
	<i>G</i> Boys		<i>G</i> Boys
E×A	.70	A×B	.70
E×B	.85	A×C	.61
E×C	.75	B×C	.73

* Significance of the tetrachoric correlations is indicated as follows:

Significantly different from zero at 1 per cent level, =====

Significantly different from zero at 5 per cent level, =====

** In order to simplify the data, *G* and *C* groups were combined for the inter-rater agreement correlations presented in Table 3. Consequently, the correlations presented here do not appear in Table 3.

Does the magnitude of the inter-rater agreement among the "experts" mean that the Rorschach protocol as a behavior sample contains communicable information concerning the subject, or does it simply mean that through intensive and similar training, the three judges have learned to a degree to *do* the same things when confronted with the test material? The former may be the case, since judges who have had little or no training are able to make similar translations of Rorschach records. Apparently several months of training and years of experience in the application of a complicated scoring system have contributed little to the agreement among the "experts" when the discriminations called for are no finer than a four-point scale.

5. *Subjective impressions of the judges*

Recorded conferences of the "experts" following their ratings made it clear that, although the raters individually were concerned with different aspects of the problem, all three agreed that it was very difficult to make over-all adjustment ratings from these Rorschachs. They felt little confidence in many of their ratings. They continually reviewed their judgments in an attempt to achieve a more normal distribution on the rating scale, yet were unable to do so. They found it difficult to rate the adjustment of these cases in terms of the norms of this group, which were available to them, because it was hard for them to depart from their preconceived notions of a "normal" Rorschach. For example, standard assumptions of normal Rorschach balance include greater use of color and a lower $F\%$ and $A\%$ than are found in this sample.

The judges agreed that the boys' groups presented quite different sorting

problems from the girls' groups. The girls' records were much richer with regard to both formal aspects and content, and the judges felt that they offered much more basis for the adjustment rating. The greater poverty of the boys' records, together with the judges' impressions of general "ugliness" of content, made it difficult for the raters to give them ratings of III or IV. In fact, on the final sort of one of the boys' groups, each judge independently failed to place any record in the "well-adjusted" category.

On the above problems, the judges were in substantial agreement. In addition, each of the raters raised particular issues which were of individual concern during the sorting process. Judge B felt it was necessary to distinguish between present adjustment and predicted future adjustment, since in many cases she would give different ratings. For instance, a girl of 18 whose Rorschach shows a tendency to relate only on a superficial level might be reasonably well adjusted in her 18-year-old peer culture, but might encounter difficulties with emotional relationships later. Judge A wanted to emphasize the necessity of different adjustment criteria for boys and girls in terms of the different cultural patterns demanded of the two sexes. For example, while fewer W than D and a $W:M$ ratio of less than 2:1 might be negative signs in a boy's record, they would not necessarily be so in a girl's test.

Many of the cues for ratings which the judges verbalized were those which have been discussed often in the literature and which have been utilized particularly in the "sign" lists of Davidson, Monroe, etc. However, the judges used them not as isolated items but as parts of a configuration which determined their weights. A few examples of other cues on which the judges were in essential agreement follow. Overemphasis on any particular blot aspect (too many W , M , Fc , etc., responses in proportion to the rest of the

record) was considered a negative sign. The use of achromatic color or white space in a colored card tended to lower the rating. A large number of *dd* responses was a much more negative sign than a large number of *dr*. Imbalance between the amount of animal movement and the use of texture (few *FM* responses and many *Fc* or many *FM* and few *Fc*) lowered the rating. Recoverability after "shock" reactions moved the rating toward the "adjusted" end of the scale. The use of a variety of content was considered a positive sign. The judges agreed that the primary use they made of the content was to indicate basic conflicts and current upsets.

Particular aspects of the Rorschachs were emphasized differentially by the three judges. Judge C was more influenced by the content than the other two raters. She was concerned with the freedom or constriction of verbalizations and seemed to use the test more like an interview than the other judges. She gave more emphasis than A or B to *M's* as a positive sign and to stereotypy and lack of color as negative signs. Thus, for Judge C, an "adjusted" Rorschach meant an "ideal" or mature record rather than a test which was average for this group.

Judge B conceived of adjustment in this group as being characterized by the popular or usual response. The absence of certain variables or responses common to this group lowered the adjustment rating. The use of the Klopfer-defined populars was especially considered, particularly the handling of the animals on Card VIII as representative of self-acceptance. Along with Judge A, Judge B felt that immaturity does not mean "maladjustment" for this group. Nor was a temporary upsetness, such as might be shown in Card VI, enough to give a "maladjusted" rating at this age.

Even more than Judge B, Judge A built her adjustment scale around the characteristics of this group. Thus, absence of color responses, stereotypy, and constriction of verbalization did not in themselves lower the adjustment ratings. However, inconsistencies in these areas, such as the absence of color-determined responses in combination with a high VIII-X%, did indicate conflict for her and consequently lowered the adjustment rating. Judge A was especially conscious of assigning differential weightings to many Rorschach variables and relationships in male and female records. For example, while *K* and *FK* were almost invariably negative signs in a boy's record, they were not necessarily negative in a girl's. Or a large number of texture-determined responses in combination with few animal-movement responses lowered a boy's rating more than a girl's.

It is interesting to note these varying approaches of the judges even though

this study does not purport to present an extensive analysis of the behavior of Rorschach interpreters. Differences among the "experts" should not be over-emphasized, however. It was as evident from the conference of judges as from the correlations and per cent agreements presented above that the raters approached their task with much the same Rorschach attitudes, and that the differences occurred primarily in the hierarchy of cues rather than in the cues themselves.

The notes made by Judge D, the "naive" rater, mentioned the following qualities as being important in determining her ratings: imagination, quality of observation, originality, ease of verbalization, affect represented, capacity to use and modify what Judge D interpreted as "normal" responses, defensiveness about the test and the concepts seen, and degree of satisfaction or dissatisfaction with responses.

Some of the spontaneous recorded comments of Judge D would be of interest to Rorschach-trained people, as they show a startling correspondence to classic Rorschach terminology and interpretive practice. For example, of a record with many good *M's*, she notes, "Enjoys her acquired knowledge. Rich inner life; reactions of self of importance to self." Of another with many white space responses, she says, "Goes out of way to see things, ignoring the close-at-hand while doing so. Sounds like belligerence; looking for trouble."

Judge E, the "novice" rater, used content a great deal, "as in a TAT." The affect expressed by the subject was important in making her ratings, as were the kinds of *M's* and color responses. Responses given to infrequently used location areas lowered her ratings.

6. Relationship of judges' ratings to Davidson Adjustment Signs (1)

In the process of making explicit their cues, the "experts" often justified the adjustment ratings they had given by listing many of the Rorschach relationships included in the adjustment "sign" lists. This was true even though the judges were aware of nonsignificant relationships between the "signs" and interview ratings of adjustment (5). They

TABLE 6
CORRELATIONS BETWEEN JUDGES' RATINGS AND THE DAVIDSON ADJUSTMENT SIGN LIST*

Judge	Boys ($N=71$)			Judge	Girls ($N=75$)		
	Sort				Sort		
	1	2	3		1	2	3
A	<u>.37</u>	<u>.39</u>	<u>.37</u>	A	<u>.45</u>	<u>.29</u>	<u>.43</u>
B	<u>.52</u>	<u>.53</u>	<u>.52</u>	B	<u>.40</u>	<u>.56</u>	<u>.37</u>
C	<u>.28</u>	<u>.31</u>	<u>.39</u>	C	<u>.36</u>	<u>.23</u>	<u>.35</u>
D	—	—	—	D	(C girls, $N=38$)	—	.12
E	(G boys, $N=41$)	—	<u>.59</u>	E	—	—	—

* Significance of these biserial correlations is indicated as follows:

Significantly different from zero at 1 per cent level,

Significantly different from zero at 5 per cent level,

did not seem to use the "signs" in an additive fashion, in which each additional "sign" present raised the adjustment rating. The context in which the "sign" occurred determined its importance as an adjustment indicator. Yet, since the "signs" were so often mentioned in the judges' conference, we were curious to know how the ratings correlated with the "sign" scoring on the same records. The correlations are presented in Table 6.

We have evidence that the "signs" are being used to some extent, since the r 's are all positive. But many other cues also seem to play a part in determining the final subjective rating. Because the Davidson list is based primarily on the formal aspects of the test, we would expect to find the highest correlations with the judges' first sort, when Rorschach content was not being used. The table shows no consistent change over the three sorts, however. The impression of Judge C, that she was least influenced by the formal aspects of the test and most influenced by the content, is confirmed here. Her correlations with the "signs" are slightly lower than A's or B's.

As would be anticipated, the Rorschach ratings of Judge D correlate low-

est of all with the formal "sign" list. The correlation for Judge E is in line with those of the "experts."

B. VALIDATION

1. The criteria

Throughout this section, the G and C groups will be discussed separately, since the outside criteria of adjustment consist of ratings made on a somewhat different basis for the two groups. The primary differences concern the intensiveness with which the cases in the two groups were studied. Both G and C groups took part in the same testing program. In addition, the G groups were much more intensively interviewed and received yearly ratings on many personality dimensions, including the scales of adjustment. For purposes of correlation with our judges' ratings from Rorschachs, two case workers who had known and interviewed the G group subjects for many years were asked to make current adjustment ratings on those cases with which they were most familiar. Their ratings were *estimates of current adjustment* on four of the scales (self-adjustment, social adjustment A, social adjustment B, and family adjustment) which had been rated at each yearly interview. Thus, the G

group case workers had available as a basis for their ratings transcripts of yearly interviews with the subjects and previous ratings on the same scales.

The interviews on which the ratings were made might be called "diagnostic," in the sense that certain basic areas were systematically sampled each year and then evaluated in codified form so that standard material was obtained on every child. Thus, the ratings were based both on behavioral information reported by the mother, the child, and the teacher, and on the observations and inferences of the case worker regarding relationships and attitudes. The coding was governed by an over-all description of the area to be coded and by an individual definition of each of the points on the five-point rating scale. These code book descriptions and definitions are presented in Appendix A.

For the C group, special current ratings were made by the case worker who had known the subjects over the entire span of the study. These ratings were made using the same definitions and codes of adjustment, but were based on more casual observation of the subjects during testing sessions at the Institute

and objective reports concerning the functioning of these individuals in the home, in the school, and on the job.

The distribution of the case workers' ratings for the four adjustment scales is given in Table 7. In contrast to the Rorschach adjustment ratings, they show a fairly normal distribution of the ratings over the five scale points. The ratings of the C-boys (previously mentioned by the Rorschach judges as the group having the most disturbed content) show some skewness with a high percentage being rated 1 or 2 on all of the scales except family adjustment. Of the C girls, 56 per cent are rated as having very good (ratings of 4 or 5) family adjustment.

Table 8 gives the intercorrelations among the case workers' ratings on the four different scales of adjustment. All of these correlations are significantly different from zero in a positive direction. There are probably two factors involved: some halo which is operating in the case worker's evaluation of each child, and a certain degree of true relationship among the various areas of adjustment. These high intercorrelations among the four ratings on each subject

TABLE 7
PERCENTAGE DISTRIBUTION OF CASE WORKERS' RATINGS OF ADJUSTMENT

Rating	Self	Soc. A	Soc. B	Family	Rating	Self	Soc. A	Soc. B	Family
<i>G Boys (N=41)</i>					<i>G Girls (N=37)</i>				
5	2%	10%	2%	7%	5	3%	8%	3%	3%
4	32%	32%	27%	24%	4	24%	22%	24%	19%
3	22%	34%	49%	41%	3	30%	43%	41%	43%
2	41%	24%	24%	20%	2	35%	19%	27%	27%
1	2%	0%	0%	7%	1	8%	8%	5%	8%
<i>C Boys (N=30)</i>					<i>C Girls (N=38)</i>				
5	13%	3%	10%	27%	5	11%	5%	5%	24%
4	20%	20%	17%	7%	4	34%	5%	13%	32%
3	20%	40%	33%	40%	3	24%	61%	42%	24%
2	33%	37%	40%	20%	2	21%	24%	37%	16%
1	13%	0%	0%	7%	1	11%	5%	3%	5%

TABLE 8

INTERCORRELATIONS AMONG THE CASE WORKERS' RATINGS ON FOUR ADJUSTMENT SCALES*

	G Boys					G Girls			
	Self	Soc. A	Soc. B	Family		Self	Soc. A	Soc. B	Family
Self					Self				
Soc. A	.60		.61	.58	Soc. A	.62		.64	.42
Soc. B			.73	.45	Soc. B		.70		.35
				.53					.49
	C Boys					C Girls			
	Self	Soc. A	Soc. B	Family		Self	Soc. A	Soc. B	Family
Self					Self				
Soc. A	.59		.75	.75	Soc. A	.71		.74	.59
Soc. B			.77	.55	Soc. B		.86		.68
				.64					.66

* Significance of these Pearsonian correlations is indicated as follows:

Significant at the .01 level of confidence, Significant at the .05 level of confidence,

will lead to few differences in correlation of the Rorschach ratings with the different areas of adjustment rated from interviews.

2. The results

a. Gross correlations with the criteria.

Figure 1 gives the Pearson correlations between the Rorschach ratings of adjustment and the case workers' ratings of adjustment for the four groups of subjects. A glance at this figure suffices to establish the most general interpretation of our results—namely, that there is no evidence of a significant positive relationship between the Rorschach ratings and the other measures of adjustment. Of the 144 correlations represented in the figure, only six are significantly different from zero at the 5 per cent level of confidence. The majority of the correlations cluster around zero, two-thirds of them in a positive direction.⁴

⁴ Between the ratings of the Rorschach "novices" and the criteria, a similar pattern of correlations prevails. Thus, the ratings by Judge E of the G boys correlate $-.01$ with the case workers' ratings of self-adjustment, $+.33$ with social adjustment A, $+.33$ with social adjust-

This low positive trend in correlations between interview-judgments and judgments based on projective techniques is a frequent finding in clinical research. It is in line with the tentative results of the research in assessing Veterans Administration trainees now being conducted at the University of Michigan (6). Guilford (3, p. 5), in discussing the extensive research in the aviation service during the war, says:

One general conclusion that was brought home to us by repeated experience is that we should have greater respect for low correlations. Tradition has taught us that unless coefficients of correlation are substantial, for example .40 or above, there is too little relationship to bother with. We must face the fact, unpleasant though it may be, that in human behavior, complex as it is, low intercorrelations of utilizable variables is the general rule and not the exception.

ment B, and $-.10$ with family adjustment. The two $+.33$ correlations are significant at the .05 level. The ratings by Judge D of the C girls correlate $+.11$ with the caseworkers' ratings of self-adjustment, $-.03$ with social adjustment A, $-.05$ with social adjustment B, and $+.03$ with family adjustment. (These correlations, based on ratings by the Rorschach "novices," are not included in Figure 1.)

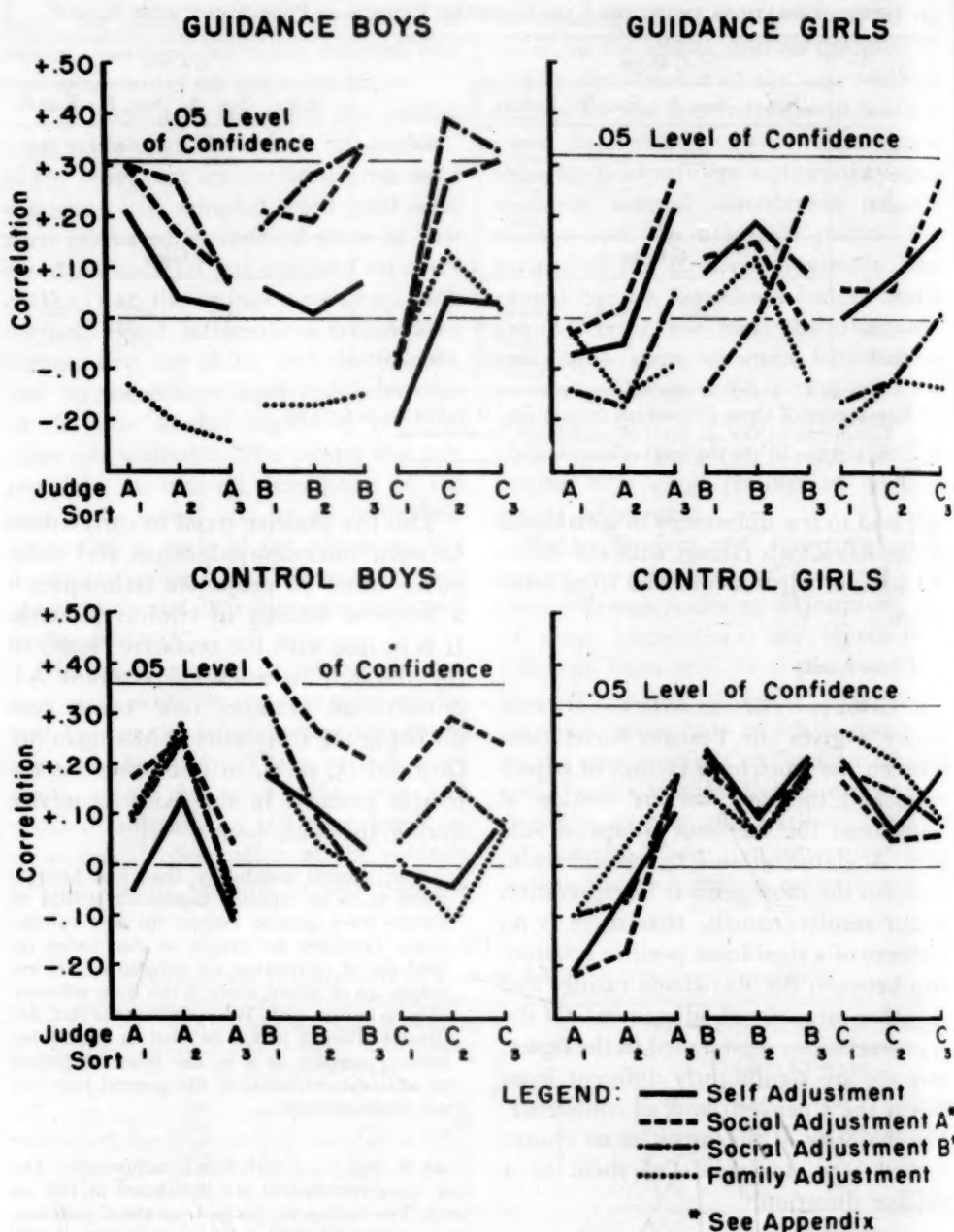


FIG. 1. Correlations of Rorschach Judges' Ratings of Adjustment with Case Workers' Ratings of Self, Social, and Family Adjustment

In evaluating our results, we are inclined to accept Guilford's bow to reality and, having made the general observation that these results may have occurred by chance, proceed, nevertheless, with one further type of analysis.

b. Analysis of variance. In line with the design of the experiment, our third general question posed in the introduction raises these specific questions: (1) Are there any differences among the judges in the degree to which their ratings correlate with the criteria? (2) Does the correlation with the criteria vary with the amount of Rorschach information available to the judges? (3) Is there a difference in the degree to which the Rorschach ratings correlate with the four areas of adjustment which constitute the criteria?

By transforming the 144 correlations represented in Figure 1 into z 's, we were able to meet the assumption of normally distributed scores and do an analysis of variance for the 36 z transformations in each of the four groups of cases (G and C boys and girls) to test the hypothesis suggested by inspection of the data, namely, that there are no significant differences in the areas enumerated above. A triple classification analysis of variance was done using the z transformations as the dependent variable. The classification variables were: the three judges; the three sorts; and the four interview scales (*viz.*, for self-adjustment, for social adjustment A , for social adjustment B , and for family adjustment). Table 9 presents the data for those effects and interactions which yielded significant F ratios.

Generalization from these results must be made with caution. Although the z in each cell is based on an N of from 30 to 41, the same persons getting each combination of treatments, it is not pos-

sible to derive a "within groups" estimate of the population variance. The use of the triple interaction variance in the divisor of the F ratio would, in this instance, seem to restrict generalization to repetition of the experiment on a sample *exactly* like ours. On the other hand, the variance of a population of z 's based on a given sample size is known, *i.e.*, the squared standard error of the z 's. Generalizations from F ratios using this population variance as the divisor, would seem to be applicable to the general population wherever significant F 's were found.

The design of the experiment is such that the z 's are not independent of each other, thus violating one of the fundamental assumptions of analysis of variance. This lack of independence has the effect of making the z 's more alike than would be the case if they were independent, thus producing smaller simple effects (*e.g.*, the between-sorts variance is small) and smaller interaction. The use of the (larger) population variance in this case would lead to the error of accepting the null hypothesis when it is not true. Consequently, we are presenting F ratios based on both the triple interaction and the population estimates of variance.

(1) *Differences among the Rorschach judges in correlations with the criteria.* Only in the C girls group does there seem to be any significant difference among the judges in the degree to which their ratings correlate with the criteria. The variance seems to be primarily a function of the fact that Judge A's correlations on the first and second sort are mostly negative, whereas the other two judges achieved low positive correlations on all three sorts. In view of the high inter-rater reliability, one would not ex-

TABLE 9
RESULTS OF ANALYSIS OF VARIANCE OF THE z 'S FROM EACH OF THE FOUR GROUPS

Source	Σ of Squares	df	Variance	F Based on Triple Interaction Variance	Level of Confidence	F Based on Population Variance	Level of Confidence
G Boys ($N=41$)							
Sorts	.041208	2	.020604	6.45	.01		
Interview Scales	.706404	3	.235468	73.75	.001	8.95	.001
Judges \times Sorts	.318733	4	.079683	25.44	.001	3.08	.05
Judges \times Interview Scales	.131260	6	.021876	7.86	.001		
Triple Interaction Variance = .0040426				Population Variance = .026316			
G Girls ($N=37$)							
Sorts	.095838	2	.047919	5.32	.05		
Interview Scales	.224327	3	.074775	8.28	.001		
Triple Interaction Variance = .0098249				Population Variance = .029412			
C Boys ($N=30$)							
Judges	.039210	2	.019605	4.51	.05		
Sorts	.091606	2	.045848	10.51	.001		
Interview Scales	.214182	3	.071394	16.41	.001		
Judges \times Sorts	.209501	4	.052375	11.33	.001		
Triple Interaction Variance = .0041409				Population Variance = .037037			
C Girls ($N=38$)							
Judges	.243016	2	.121508	56.11	.001	4.28	.05
Sorts	.071579	2	.035789	16.52	.001		
Judges \times Sorts	.208117	4	.052029	23.90	.001		
Triple Interaction Variance = .0028245				Population Variance = .028571			

pect much difference among the judges in the matter of validity.

(2) *Differences among the three sorts in correlations with the criteria.* Variance from one sort to another is significant in all four groups. Obviously, however, it is not in a consistent direction. The correlation with the criteria may drop with one judge from the first to the third sort and with another increase, while these same judges' correlations will reverse their direction in another group. Increasing amounts of Rorschach information, then, did not result in increasing (or decreasing) correlation with the criteria.

The only interaction which is significant for three of the four groups of

subjects is that of judges and sorts. In other words, the judges are acting differently on the different sorts in terms of their correlation with the criteria. That this is the case is obvious from inspection of Figure 1.

(3) *Differences among the interview scales in correlations with the Rorschach ratings.* There is one F significant at the .001 level of confidence when we use the population variance as the divisor. This is variance among the correlations with the four different interview scales for the G boys. The majority of the G boys were rated by one case worker who apparently (see Table 8) made greater discriminations among the four types of adjustment than the other raters. Using the triple

interaction variance in the F ratio, the correlations vary significantly among the four interview scales for all groups except the C girls. An inspection of Figure 1 will show that the judges' Rorschach ratings correlate more highly with the case workers' ratings on the scales of social adjustment than with ratings on the other adjustment scales. All of the significant correlations are with the two social adjustment scales, while the majority of the correlations with family adjustment are low negative.

Combining the impressions from all of the data, one is led to the conclusion that perhaps the most interesting aspect of the study is not the subjects, their Rorschachs, and their case histories, but the workers themselves, the Rorschach judges, "expert" and "naive," and the case workers. Had we limited our study to G boys only, for example, our conclusions might have been clearer and more confident than they are when we check them with results from the other three groups also. One major difference among the groups arises from the case workers' ratings of the subjects. For example, the mean interscale correlation for the worker who rated most of the G boys was .50, whereas the mean for the worker who rated all of the C cases was .70. This fact alone, since the Rorschach judges rated on only one scale, would influence the difference in correlation with the four scales which shows up in Figure 1.

c. *Case studies.* In a further attempt to explore the relationship between the two types of ratings, some individual cases with wide discrepancies were studied. There were seven cases whose Rorschach ratings were in the fourth quartile and whose interview ratings were in the first quartile. There were six cases whose

Rorschach ratings were in the first quartile and interview ratings were in the fourth. The histories of those discrepant cases from the G group were thoroughly perused without any hypotheses in mind, in the hope that they might suggest hypotheses which could be checked against the cases from the C group. To a certain extent (even more than one might expect with such a small number of cases), this was successful.

Five of the seven cases whose Rorschach ratings were much higher than the case workers' ratings had very similar family patterns. These included desperately severe economic insecurity during childhood, much marital conflict between the parents with threatened separation and divorce, and siblings who, under these conditions, became delinquent. In all of these cases our subjects, although they certainly had problems and tensions, might be described as the best adjusted members of very poorly adjusted families. Perhaps their "good" Rorschachs reflect the capacity to make some sort of adjustment to an almost impossible environment. The case workers' ratings, on the other hand, used the whole study group as a reference population, i.e., the subjects were rated on their adjustment relative to the other subjects in the group, *not* relative to the job of adjusting they had to do. It is interesting too that absorbing interests in music or religion or both were common to all of these subjects and may have been important factors in their adjustment.

The pattern in the six cases whose interview ratings were high and Rorschach ratings low is not quite so clear. Three of these subjects have in common a high socioeconomic status, which has resulted in smoother life circumstances for them than for the children in the

other group. There is no doubt that they are attractive young people, socially successful in a school system with definite social hierarchies. The conflicts resulting in a "poor" Rorschach have perhaps not been precipitated and resolved in this smooth, non-tension-producing environment. Another two cases are quiet and unanimated, with relatively low IQ's for our group and a low level of aspiration; they have accepted their personal and social liabilities and appear to the case worker on these grounds to be adjusted. To the Rorschach worker, however, real adjustment is questioned because their Rorschachs seem flat and withdrawn.

V. SUMMARY

The Rorschach records of 146 normal 18-year-olds (71 boys and 75 girls) were rated for over-all adjustment on a four-point scale by three Rorschach "experts." Each of the records was rated three times by each of the raters. In the first rating or "sort" of the subjects, only the formal scoring categories of the Rorschach were used; in the second rating or sort, the distribution of these categories over the ten cards was included; in the final rating or third sort, the whole protocol was used. Data from the three rating trials or "sorts" were separately treated in this study.

The ratings were made with the subjects divided into four groups: 41 of the boys (*G* boys) and 37 of the girls (*G* girls) had been intensively studied throughout their lifetimes as members of the Guidance Study of the Institute of Child Welfare, University of California. The remaining 30 boys (*C* boys) and 38 girls (*C* girls) were members of a matched control group in the Guidance Study and had been studied less intensively over the same period of time. For purposes of

validating the Rorschach ratings against ratings of adjustment made by case workers, these four groups were treated separately, since the amount of information available to the case worker as a basis for adjustment ratings differed in the *G* and *C* groups.

The Rorschach judges consistently placed from 61 to 71 per cent of the cases on the "maladjusted" end of the four-point scale. However, the magnitude of the inter-rater agreement correlations showed that "experts" can agree, not only among themselves but also with judges having little or no Rorschach training. (Two other judges, one of whom had only a few months of Rorschach training and one of whom had absolutely none, rated a small group of records.) In verbalizing their cues, the "experts" mentioned many of the formal Davidson adjustment signs (their ratings correlated $+ .23$ to $+ .56$ with the Davidson list), but indicated that the signs were weighted with regard to the total context and that many other less objective cues were also used.

Concerning inter-rater agreement, there was no particular advantage or disadvantage to using the full Rorschach protocol over the formal aspects alone. With the addition of content, the raters changed more of the girls' ratings than the boys'. There was more consistent inter-rater agreement on boys' records than on girls' within a single sort, and more consistent intra-rater agreement on boys' records than on girls' between sorts. More boys than girls were rated as "maladjusted." This is in line with the results of our analysis of "neurotic" signs (5), in which the boys consistently received higher scores, and the analysis of adjustment signs (5), in which the boys received lower scores.

Correlations of the Rorschach ratings with the case workers' adjustment ratings were primarily low positive. The lack of a statistically significant relationship can lead only to the conclusion that the Rorschach, used blindly (i.e., without supplementary background and observational data) is of little value in predicting adjustment ratings made primarily on the basis of interviews and case histories.

Within this framework of no validating relationship, however, a few interesting trends emerge. Those Rorschach ratings which did correlate significantly with the criteria were with ratings of social adjustment. The correlations with family adjustment tended to be negative. Discrepancies among the results with the different workers point up the difficulty of using clinical judgments as validating criteria. A few individual cases with wide discrepancies between the two types of ratings were examined and interesting personal and sociological concomitancies observed. These concomitancies are worthy of verification by further investigation. As yet, however, the sources of the discrepancies between the Rorschach ratings of adjustment and the interview ratings remain undetermined. It is suggested that differing social background of the subjects may be responsible for some of the discrepancies; sociological biases of the case workers may also be influential; and finally, of course, there is the possibility of inadequacies both in the Rorschach and the interview data, as such.

VI. CONCLUSIONS

After hearing of the results of this study, an academic psychologist said to the authors, "Now are you finally ready to give up the Rorschach?" That, of course, is too simple. We know that a

Rorschach interpretation can add meaning to the understanding of an individual case. But it is true that our confidence in the ability of Rorschach workers to analyze records of normal subjects for use in group research has suffered.

The authors would like to suggest five possible sources of the negative results in our validation attempts, only one of which is the Rorschach itself.

1. One of the difficulties may lie in the nature of our sample. It is primarily a normal sample with no pathological extremes, and finer discriminations than are obtainable by a brief four-point scale are necessary to differentiate homogeneous samples.

2. The Rorschach may represent such a limited behavior sample, so sensitive to temporary disturbances, that it is not useful for predicting "adjustment" in other situations or at other times. Or the test may pick up such evidence of instability as exists in the individual without ever tapping his areas of stability. The Rorschach may offer us no picture of present adjustment but rather predictions which only time and situational concomitancies can test. This projective technique may be one which will give us knowledge only of the subject's inner potentialities and functioning, never of his overt behavior. Another point which arises involves the theoretical question of the function of fantasy. If one subscribes to the theory that fantasy offers a "safety valve" device for many individuals, then, of course, we would expect no one-to-one relationship between the pictures of adjustment presented in the overt and covert areas.

3. The third possible source of our negative results is the Rorschach rater. A large part of Rorschach training consists of learning to recognize signs of abnormality which have been established

on pathological samples, with little attempted comparison with randomly selected normal groups. Our raters have in common with other Rorschach workers this background of training and vicarious experience through the literature, which has sensitized them to the conventional indicators of maladjustment. Evidence of this concern is to be found in the distribution of their ratings in which the majority of cases were rated "maladjusted." However, the positive correlations of our "naive" rater with the other judges and her zero correlations with the validating criteria indicate that the problem does not lie completely in stereotyped thinking and training on the part of our "experts."

4. A fourth series of problems concerns the criterion variables used, namely, case workers' ratings of adjustment. Although all case workers used the same rating scales, they were often of different theoretical orientation. While all had access to previous case records, they varied with regard to the length of time they had known the subjects, the degree of openness of the relationship, and thus with regard to their understanding of the dynamics of the cases. The difficulty posed by differences between the G and C groups with regard to intensiveness of knowledge about the cases further complicates our attempts to pin down the factors involved in our negative results.

5. A fifth possible factor in the interpretation of our results is the semantic difficulty inherent in the use of the concept of adjustment. The concept of overall adjustment, used by the Rorschach raters, is too broad and too vaguely defined. The meaningfulness of so inclusive a concept of adjustment, without regard to time, place, or area of functioning, may be questioned. Further, the concept of adjustment used by the case workers was specified as to these variables, while that used by the Rorschach raters was not.

These considerations have some implications for future validation attempts with the Rorschach. Research clinicians using this tool might well spend more time and effort collecting and studying the records of adequately functioning individuals, in order to counteract their biases in the direction of abnormality. The results also suggest that narrower, much more specific areas of personality be studied through the Rorschach for ease in finding suitable validating criteria. Another point, which has been made many times before, might be reiterated here. Much more meaningful results may be obtained when we set up our experiments so that the Rorschach is not tested in isolation, but rather investigated in terms of the additional predictive significance which it gives in combination with other clinical measures.

APPENDIX A

Adjustment to self (size, sex, physical make-up, abilities, disposition, past misdemeanors, etc.)

- (1). Extreme lack of confidence shown either through excessive self-depreciation, unfavorable comparison, or marked braggadocio. Intense guilt feelings, or feelings of inadequacy.
- (2). Not as marked as 1 but marked unacceptance of one or two items of make-up. However,

if to the point of dominating whole personality, classify under 1. Or many mild points of discomfort. Over-adjusts to others at own expense.

- (3). Lacks confidence in some particulars but for the most part is unself-conscious or unaware of assets or liabilities. Occasional tension or conflict about self but not characteristic.
- (4). Recognizes handicaps but accepts them real-

istically and without much tension. Little evidence of compensatory behavior.

- (5). Full of confidence—accepts limitations and assets easily.

Adjustment to family (codings on the severity of tension—whether it involves one or more family members, including step-parents. Tension-producing relationships whether shown openly or by withdrawing or in compensatory demonstrativeness or such intense attachments or such ingrown relationships that normal, friendly relations to outsiders are precluded).

- (1). Marked hostility or insecurity with parents, feels he is not loved, discriminated against, or fails to meet parents' expectations or sibs' accomplishments. Humiliation over parents' faults, occupation, income, etc. Child subjected to rather intense, ambivalent swings. Practically no compensatory satisfaction.
- (2). Same trends, but not so extreme as in 1. Much more strain than average but some sources of security. Very poor relationship with one parent. If this poor relationship is with a parent with whom he seldom comes in contact, it should be included here *unless* it dominates his other family feelings, in which case the coding is 1.
- (3). For the most part, low-grade tension; not markedly unsatisfactory, but enough insecurity in relationship to family members to make for strain or chronic low-grade embarrassment over family members; mild jealousies.
- (4). Absence of real strain—occasional episodes of insecurity or lack of outstanding satisfactions.
- (5). Complete security in family affection, parents, sibs. Genuinely proud of parents or a realistic (unembarrassed) acceptance of parents' faults, limitations, or assets. No evidence of tension-producing attachment.

Social adjustment

- A. (Refers, as used here, to social techniques, ease with which a child evokes acceptance, approval, friendliness or loyalty; or incurs disapproval, distrust, or hostility. Emphasis is on a social criterion rather than on child's security in social situations—this last being handled under such items as shyness, friendliness, quarrelsomeness, etc.)

- (1). Extremely unsuccessful in social relations. Intensely disliked or resented by practically all of his playmates. Very poor social techniques and constantly does things which offend his social group. Actively avoided.
 - (2). Unsuccessful in social relationships; not popular but does not give rise to the intense reactions of 1. Ignored or not spontaneously included in group play, etc. Does things which annoy. Definitely disliked by many children; has some supporters.
 - (3). Not outstanding. Normally successful in having people like him. Some playmates may not like him but has approving supporters. Or successful with one sex but unsuccessful with the other.
 - (4). More popular than the average; has good social techniques.
 - (5). Unusually successful in social relations. Inspires friendliness or loyalty. Has many people devoted to him. Constantly gives rise to such statements as, "Isn't he a swell guy?" "She's a grand person," etc.
- B. (Refers, as used here, to child's feeling of tension or confidence and satisfaction in social relations)
- (1). Acutely uncomfortable, dislikes people, feels he is treated unfairly or is disliked by others. Chronically sulks or is aloof.
 - (2). Definite but less acute tension than 1. Feels he is not liked as well as other children. "Children don't play with me much." Or worried because one or two children dislike him. Insecurity may take form of shyness or aggressive attempts at social overtures.
 - (3). Takes his social relationships for granted, doesn't occur to him to consider whether he is liked or disliked. Some frictions but "have as many friends as most children do." May have few friends but apparently has no tension or worry on this score. Feels secure with one sex but a bit uneasy with the other.
 - (4). More confidence in social relationships than average. Very little tension and many satisfactions.
 - (5). Likes children very much; enjoys being with them. Is enthusiastic and happy in relations. Confident, feels socially secure.

BIBLIOGRAPHY

1. DAVIDSON, HELEN H. *Personality and economic background*. New York: King's Crown Press, 1943.
2. DUNLAP, J. W., AND McNAMARA, W. J. A graphical method for computing the standard error of biserial r . *J. exp. Educ.*, 1934, 2, 274-277.
3. GUILFORD, J. P. Some lessons from aviation psychology. *Amer. Psychologist*, 1948, 3, 3-12.
4. HAYES, S. P., JR. Tables of the standard error of tetrachoric correlation coefficient. *Psychometrika*, 1943, 8, 193-203.
5. IVES, VIRGINIA, GRANT, MARGUERITE Q., AND RANZONI, JANE H. The "neurotic" Rorschachs of normal adolescents. *J. genet. Psychol.*, 1952.
6. KELLY, E. L., AND FISKE, D. W. The prediction of success in the VA training program in clinical psychology. *Amer. Psychologist*, 1950, 5, 395-406.
7. KLOPPER, B., AND KELLEY, D. M. *The Rorschach technique*. Yonkers, N.Y.: World Book Co., 1942.
8. MACFARLANE, JEAN W. Studies in child guidance. I. Methodology of data collection and organization. *Monogr. Soc. Res. Child Develpm.*, 1938, 3, No. 6.
9. MCFATE, MARGUERITE Q., AND ORR, FRANCES G. Through adolescence with the Rorschach. *Rorschach Res. Exch.*, 1949, 13, 302-319.